Tumkur University

Department of Studies and Research in Library and Information Science

B.H Road, Tumkur-572103

Synopsis on

Mapping of Web Citations: Pertinacity and Decadence of URL Citations in Civil Engineering Scholarly Publications.

Submitted by

Shanthakumari K
Research Scholar
(Reg. No. **18LI003)**
Department of Studies and Research in Library and Information Science
Tumkur University
Tumkur

Submitted to

Dr. Keshava
Professor
Department of Studies and Research in Library and Information Science
Tumkur University
Tumkur

2019

1. Introduction

Among all networking technologies, the internet technology is most useful, powerful and economically available in the present era of information communication technology. The internet gives access to an overwhelming amount of information, data and electronic information services at a finger tips (PSG Kumar ,2014).

The internet has revolutionised our society. The internet has created avenue that is bringing people of the world as well as information resources together. The internet and World Wide Web have created a avenue for the coherent communication and dissemination of information in digital forms across the world. A document or message sent through internet and available over World Wide Web may contain texts, images, graphics or any other form of information. The World Wide Web has no limitation. It is unstinting and embody the whole spectrum of knowledge available elsewhere in this world and possibly besides the cosmic limits of our planet.

The acclamation of web among academicians at one hand the increased amount of information available on the web on the other hand has persuade the use of web citations by authors of scholarly world (Spinellis, 2003). Web citations have been proposed as complements to, even replacements for, traditional citations (Liwen Vaughan and Debora Shaw, 2005) Web citations have become the norm in scholarly literature (Goh & Ng, 2007). The web citations cited in research work should be obtainable. The web citations cited in a research work might be a guiding tool for further research.

Therefore, availability of cited content over time, will help authors, editors and publishers develop policies and conventions that will ensure long-term access to cited web content (Amirhosein Mardani, 2012)

Internet is dynamic in nature. The advent of web has altered the dynamics of information access in that information is no longer restricted to a physical presentation medium, no is it reliant on geographical location. Any internet linked computer allows researchers to access information that may have been previously unavailable to them (Sellitto, 2005).

The online sources are amenable for correction. They can be overwritten and in many cases, their authors are not committed to long term storage and maintenance (Casserly & Bird, 2003). This has lead to the problem of online information decay. The permanency of these cited web resources in research work has been questioned by many authors (Rumsey, 2002; Spinellis, 2003; Germaine, 2000; Lawrence, 2001) with disappearance rates of cited URL. Hence, the disappearing act of web resources leads to create a problem. The issue of URL permanency has become important due to the way that Web-located information is being increasingly cited in both general and academic publications. When an author cites a resource located on the Web there is fundamental assumption of resource permanency, that is, the particular information resources will be found at the cited location. However, given that Web located resources are being increasingly cited, there has been concern on the way that URL references are disappearing (Rumsey, 2002).

Therefore, the web needs preservation initiatives to save the contemporary knowledge which is the secondary sources for the research community. Web archiving initiative collects; preserve the web resource for posterity. The same is accessible for longer durations. The web archiving initiatives has begun since 1996, when four major initiatives have started.

Web archiving is a procedure for controlling the URL decay and thereby the loss of web resources. Many web archiving initiatives have been launched by

developed countries. Few among them allow the full text search for archived materials and provide URL history. The scope ranges from regional to international level. They are the paramount tools to recover missing web resources. Archived web information has become a benevolence to the academic and research community. In this context, the present study has to be conducted to estimate the Pertinacity and Resumption of URL Citations Used in Civil Engineering Scholarly Publications through the selected web archiving initiatives. This study also focuses on the capabilities of web archives to recover the missing web resources. A comparison is going to be made between Internet Archive and Pandora-Australia's web Archive to know their capabilities.

2. Statement of the problem

The statement of the problem of the present study is *"Mapping of Web Citations: A Contemplative Study of Pertinacity and Decadence of URL Citations Used in Civil Engineering Scholarly Communications"*

3. Objectives of the study

1. To find out growth rate and doubling time of civil engineering scholarly publication
2. To determine the authorship pattern and collaborative measures of civil engineering scholarly publication.
3. To know the use of URLs as Citations cited in Civil Engineering scholarly publication.
4. To measure the rate of decadence of URLs cited in Civil Engineering scholarly publication.
5. To determine the half-life period of URLs cited in Civil Engineering scholarly publication.

6. To compare the efficiency of Internet Archive and Pandora-web archive in recovery of vanished URLs.
7. To suggest the possible solutions to avoid decay of URLs.

4. Need of the study

The web has vouched the eye-opening production of content. The web pages are having different authorities such as authors, individuals, and publishers. Though web pages have different authorities, they should assure that the information over the web has to be existed for longer duration. But the nature of web is potent. Because the potent nature of the World Wide Web offers initial ease of access but unreliable subsequent retrieval, and with decay rates increasing over time, the internet may prove to be an inhospitable medium, especially web based research(Dimitrova and Bugeja, 2015) [The half-life of internet references cited in communication journals.] Therefore, web needs sustain initiatives to fight web decay. Here, a question has risen that whether the ratio of information production on the web and their archiving go parallel. Statistics from previous studies has shown 42 web archiving initiatives have been created since 1996 so far.

Web archives now exceed 2PBs of data and encompass over 150 billion captures collected from 1996 to the present culled from every domain, over 200 million web sites and 40+ languages. This archival database expands with the Internet, and so grows by nearly 100TBs every month. Usage of Internet Archive's web collections via the Wayback machine average 400-500 requests per second.

But the production of web resources on the web is comparatively high than the web resources that are being archived by web archiving initiatives. It is estimated that the Indexed Web contains at least 2.64 billion web pages (WorldWideWebSize.com) in

its web. It is important to know which web archiving initiatives is best to serve the purpose of identifying missing web resources.

Therefore, it is imminent to understand the scope and coverage of web archiving initiatives. Therefore, this study is an attempt to know the efficiency of selected web archiving initiatives in retrieving missing web resources. A thorough literature review also revealed that no comprehensive study has been made to check the capabilities of web archiving initiatives. Therefore, a long term study would be helpful to know the efficiency of web archives which are useful tools to retrieve archived data.

5. Review of Literature

Use of web citations

Web is fated source of information. The use of web is common among academic and research community. The use of web played an important role in scholarly world. Authors of scholarly literature in many disciplines are banking upon the Web for information. The studies carried out to know the use of web citations has shown the fact.

Lawrence et al. in 2001 studied 270,977 computer science journal papers, conference papers, and technical reports. They had extracted 67,577 URL citations from 100,826 cited works. Rumsey (2002) examined the use of URL citations in law review articles. This study witnessed an increased growth of URL citations. The law review articles citing to the Web had increased from 130 (0.57%) articles in 1995 to 5462 (23%) in 2000. Another study by Bhat and Sampath Kumar (2008) studied nine open sources e-journals published and found 11,199 web citations out of 25730 overall citations. Recent studies carried out by many authors such as Janakiramaiah and Doraswamy, 2011- (36.55%); Mardani, 2012- (24.90%); Sampath Kumar and

Manoj Kumar, 2012 (18.77%) have shown that the use of web citations in scholarly literature is a common phenomenon.

Web Decay

McCown et al. (2005) had discovered 28% of URL citations cited in D-Lib Magazine were failed to recover out of 4387 URLs referenced in 453 articles published from July 1995 to August 2004. Their last check showed the inconsistency rate had increased to 30%. Dimitrova and Bugeja (2007a) studied the URL citations of five leading journals (2000-2003) in Journalism and Communication and found 1126 URL reference. Their results showed that 39% of web citations were inaccessible in 2004. Goh and Ng (2007) investigated the web decay in three leading Information Science journals. They have found that 31% of all citations were not accessible during the time of testing.

Ducut et al. (2008) conducted a survey on MEDLINE URL's and found out of 10,208 URLs 2245 were not accessible with the percentage of 21.99%. In recent year Nagaraja et al. (2011) examined persistence of 2503 web references from 1133 articles published from 2005 to 2007 in PLoS Medicine. 17% of the URL citations were defunct and the rate of URL attrition increased as time elapsed. Thus it is clear from the above review that the web decay is common phenomena and it is major.

Recovery of missing URLs

Although web decay is a major problem, web archiving initiatives have provided a provision to recover missing web resources. Many studies were attempted to recovery missing URLs through Wayback machine and Google search engine. In the year 2003, Casserly and Bird conducted a study on the web citation availability in the articles published in library and Information Science journals in during 1999 and

2000. Researcher found that 81.4% were available on the web, and searching the Internet Archive increased the availability rate to 89.2%.

Dimitrova and Bugeja (2007b) used Wayback machine and Google to retrieve missing online citations and found Wayback machine was more efficient than the search engine. They had yielded 53.4% of websites through wayback machine. Russell and Kane (2008) had used wayback machine and retrieved 52 missing URLs out of 510 cited URL citations. They have identified that the use of wayback machine had reduced the rate of missing URLs from 91 to 39.

A study by Tajeddini et al. (2011) explores availability and decay of URLs cited in articles of six Library and Information Science (LIS) journals published by Emerald, Science Direct and Sage. Research findings indicated that 66% of articles had web citations. Original accessibility of web citations was 66% which improved to 95% by second check availability using the Wayback machine and the Google. The study of above literature has shown that the way back machine is an effective instrument to archive the web pages. These web archives allow in retrieving the missing URL citations cited in scholarly literature.

6. Scope and limitations of the study

This study would explore the trends and decay of the web citations which were cited in the Civil Engineering scholarly literature. Present study will cover the following journals of reputed publishers for the period of 10 years i.e. 2011-2020. These UGC approved journals have been selected for the study on the basis of their impact factors and reputation.

Present study is an effort to check the efficiency of two web archiving initiatives viz., Internet Archive and Pandora. The scope of the study is limited to these two web archiving initiatives only

SELECTED LIST OF UGC APPROVED CIVIL ENGINEERING JOURNALS					
Sl. No	**Name of the Journal**	**Publisher**	**Frequency**	**Impact Factor**	**Country**
1	Proceedings of the Institution of Civil Engineers: Engineering Sustainability	Institute of Civil Engineers	Bimonthly	1.5 (two year)	UK, London
2	Proceedings of the Institution of Civil Engineers: Urban Design and Planning	Ice Publishing	Bimonthly	SJR 0.238	Scotland
3	Road Materials and Pavement Design	Taylor & Francis	10 Issues per year	2.199	UK
4	International Journal of Pavement Engineering	Taylor & Francis	Monthly	2.322	UK
5	International Journal of Structural Engineering	Inderscience Enterprises Ltd.	Quarterly		UK
6	ISH Journal of Hydraulic Engineering	Taylor & Francis	3 Issues per year	0.111	UK
7	International Journal of Sustainable Transportation	Taylor & Francis	10 Issues per year	1.892	UK
8	Building Research and Information	Taylor & Francis	8 Issues per year	3.468	UK
9	Journal of Construction in Developing Countries	Penerbit Universiti Sains Malaysia	Biannually		Malaysia
10	Mechanics Based Design of Structures and Machines	Taylor & Francis	Bimonthly	1.115	UK
11	Journal of Transportation Engineering	Asce-American Soc Civil Engineers	Monthly	(SJR) 0.59	Baltimore US
12	Journal of Composites for Construction	Asce-American Soc Civil Engineers	Bimonthly	(SJR) 1.84	Baltimore US

7. Methodology

Selection of articles and URL citations

Articles published in 10 journals in Civil Engineering scholarly publication will be considered for the present study. The time frame for the analysis will be the publication years 2011–2020 i.e. 10 years. All research articles published during the 10 years period will be downloaded and saved. The references that appeared as a list

at the end of the article under the bibliography or reference section will be considered for the study.

Testing of URLs

The study will attempts to check the availability of URL citations and their persistence, W3C link checker (http://validator.w3.org/checklink) will be used to check the URLs existence (Sellitto, 2005). After initial check, the checked citations will be grouped as 'active' and 'missing' URL citations.

Recovery of missing URLs

The study aims to recover vanished URLs which would found missing during the initial check using W3C link checker. The Internet Archive and Pandora web archives will be used to recover the vanished URLs. The study also will be aims to compare the efficiency of the Internet Archive and Pandora web archive in retrieving the vanished URLs.

8. Chapter scheme

I. Introduction

This chapter covers details such as introduction to the topic, statement of the problem, objectives of the study, need of the study, review of related literature, methodology used in the study and chapter scheme.

II. Internet Archive and Pandora: A Profile

This chapter will give an overview of the two web archive initiatives i.e. Internet Archive and Pandora- Australia's Web Archive.

III. Literature review

This chapter will be an attempt to review the literature related to the research topic.

IV. Analysis and Interpretation of data

This chapter covers the data collected regarding the use and decay of URLs, recovery of vanished URLs through Internet Archive and Pandora web archive. The data will be presented in the form of tables and graphs.

V. Findings, Recommendations and Conclusion

The analysis of the collected data will yield useful findings. Recommendations would be drawn on the basis of those findings. This chapter would be an attempt to present the findings and recommendations of the study.

References

Bar-Ilan, J., & Peritz, B. C. (2004). Evaluation, continuity and disappearance of documents on a specific topic on the web: A longitudinal study of informetrics. *Journal of the American society for Information Science and Technology*, 55(11), 980–990.

Casserly, M. F., & Bird, J. E. (2003). Web citation availability: Analysis and implications for scholarship. *College and Research Libraries*, 64(7), 300–317.

Dimitrova, D. V., & Bugeja, M. (2007a). The half-life of Internet references cited in communication journals. *New Media & Society*, 9(9), 811–826.

Dimitrova, D. V., & Bugeja, M. (2007b). Raising the dead: Recovery of decayed online citations. *American Communication Journal*, 9(2)

Goh, D. H. L., & Ng, P. K. (2007b). Link decay in leading information science journals. *Journal of the American Society for Information Science and Technology*, 58(1), 15–24.

http://pandora.nla.gov.au/statistics.html

Janakiramaiah, M., & Doraswamy, M. (2011). Measuring impact of web resources in conference proceedings: A citation analysis. *In CALIBER 2011* (pp. 541–549).

Kumar, PSG (2014). Changing dimensions of library and information science. New Delhi: Ess Ess

Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, 4000, 107–109.

Mardani, A. (2012). An investigation of the web citations in Iran's chemistry articles in *SCI. Library Review*, 61(1), 18–29.

McCown, F., Chan, S., Nelson, M. L., & Bollen, J. (2005). The availability and persistence of web references in *D-Lib Magazine*. Available from http://iwaw.europarchive.org/05/papers/iwaw05-mccown1.pdf

Nagaraja, A., Joseph, S. A., Polen, H. H., & Clauson, K. A. (2011). Disappearing act: Persistence and attrition of uniform resource locators (URLs) in an open access medical journal. *Program: Electronic Library and Information Systems*, 45(1), 98–106.

Rumsey, M. (2002). Runaway train: Problems of permanence, accessibility, and stability in the use of web sources in law review citations. *Law Library Journal*, 94(2), 27–39.

Sampath Kumar, B. T. and Manoj Kumar, K. S. (2012). Persistence and half-life of URL citations cited in LIS open access journals. *Aslib Proceedings: New Information Perspectives*, 64(4), 405-422.

Sellitto, C. (2005). The impact of impermanent web located citations: A study of 123 scholarly conference publications. *Journal of the American Society for Information Science and Technology*, 56(7), 695–703.

Spinellis, D. (2003). The decay and failures of web references. *Communications of the ACM*, 46(1), 71–77.

Veena, B., & Sampath Kumar, B. T. (2008). Web citation behaviour in scholarly electronic journals in the field of Library and Information Science. *Webology*, 5(2). Available from. http://www.webology.ir/2008/v5n2/a57.html

(Shanthakumari K) **(Dr. Keshava)**

Signature of the Candidate **Signature of the Guide**